Financial Research Advisory Committee
Data and Technology Subcommittee

**Proposal for an**
**Ontology Viability Project**
July 2014

The OFR has statutory mandates to improve the quality and scope of financial data and to "prepare and publish" a financial instrument database. Developing and implementing data standards and the conceptual framework to organize financial data will promote achieving those related mandates.

Two OFR initiatives will further those efforts. First, the OFR is working with the Commodity Futures Trading Commission (CFTC) to improve the quality of data on derivatives transactions that are being provided through swap data repositories (SDRs), as mandated by the Dodd-Frank Act. In addition, the OFR is exploring ways to evaluate the recommendation of the Financial Research Advisory Committee (FRAC) to consider a comprehensive ontology for financial instruments as part of the overall financial instrument database effort.

These two efforts are related. The challenges facing SDRs are, in critical ways, a particular version of the the broader problem the OFR faces regarding its obligation to publish an instrument database. Strategies that are helpful in improving the data provided through the SDRs should help inform the OFR's choices regarding its instrument database mandate.

Over the last several years private sector groups have been working to develop an operational ontology for financial instruments, business entities and associated analytic tools. An ontology is just a set of agreed upon definitions of the key characteristics of financial contracts—the organizing framework that incorporates the data standards into a coherent whole, and helps relate instruments to each other. That work is already well advanced for interest rate swaps and business entities (counterparties). This represents an important opportunity for the OFR. We believe that the existing ontology for interest rate swaps and counterparties can contribute to improving the data provided through SDRs and tell us a lot about what ontologies can, and can not, contribute to the OFR's broader data standards mandate.

The Financial Research Advisory Committee recommends that the OFR launch a "proof-of-concept" project to evaluate how this standard ontology for interest rate swaps and business entities can contribute to the work that the OFR is doing with the CFTC to improve the data provided through the SDRs. If this ontology project proves useful in this context, this should be a good indication of the practicality of using ontologies for the OFR's wider objective of making financial data more coherent, including its mandate to publish an Instrument Database.

Market participants are providing a tremendous amount of information to the SDRs. The first order problem is to get the most out of that information. The first step in the proposed project would be to align the information that is already being provided to SDRs to an existing ontology. That is, to associate a validated set of precise definitions to the data fields that are already being provided to the SDRs.

If that can be done it will be possible to better assess what sorts of aggregate information can, and can not, be efficiently and accurately generated from the data that is already being provided to the SDRs. Note that flexible analytic techniques can be applied to data that is structured in a standard ontology language and that should make it easier and far more efficient to generate aggregate statistics and to perform flexible queries on the data.

In order to meet the need for consistent, meaningful, and high-quality data on derivative positions it is likely that greater degree of harmonization in the data collected by different SDRs will be needed. One of the primary objectives of the project should be to assess whether or not existing ontologies can be the basis for consistent and adaptable standards for the data collected by SDRs. This project can contribute not only to the OFR's work with the CFTC but also international initiatives in this area such as the Financial Stability Board's Aggregation Feasibility Study Group's work on indentifying strategies for aggregating data OTC deritvatives from trade repositories.

It is also worth noting that attaching precise definitions to the data that is being provided SDRs will also make it easier to carry out data quality checks to help identify inconsistencies and errors in the data that is being provided by market participants to the SDRs.

Doing the analysis to understand what can, and can not, be done with the existing information that is being provided to the SDRs should also help to identify what financial participants need to do to align their internal repositories to common meaning in the data that is reported to the the SDRs. Data standards are an essential element of successful data management systems maintained by individual financial firms. This principle has been recognized by international supervisors and codified in the Basel Committee's "Principles for effective risk data aggregation and risk reporting."[1] A number of financial firms are already aligning their internal data to industry-standard ontologies.

Of course ontologies are not a panacea for all the challenges relating to the data collected by the SDRs. In particular, the development and adoption of robust identifiers for individual transactions will also be critical.

We suggest that the proposed project focus on interest rate (IR) swaps. The focus on IR swaps should help the OFR to develop a viable approach for solving existing data challenges associated with an important set of derivative transactions. We believe that the project should seek to test the value of ontology in addressing the SDR data challenges in the following specific ways:

- Aggregation of data from multiple repositories within and across a broad spectrum of financial institutions

- Classification of the full spectrum of IR swaps into product categories based on the facts about the instrument and align them with existing reporting regimes and taxonomies such as FpML,

[1] See Basel Committee on Banking and Supervision, "Principles for effective risk data aggregation and risk reporting," January 2013.

FIXML and ISO 20022

- Validation that the data being reported about various IR swap types are both complete and structurally accurate across the complete reporting lifecycle (i.e. from internal participant repository → transaction → SDR → reporting to regulatory agency)

- Performing flexible risk queries about such areas as: business entity ownership/control structure; transitive relationships among multiple business entities; concentration and exposure based on role, cash flow, instrument type, asset class, position or country; and the measurement of centrality

- Ease and cost-effectiveness of integrating ontology into existing technology environments

- The generation of reports and visual diagrams to support complex analysis

In order to demonstrate the viability of ontology and related analytic tools in a "real world" environment we strongly suggest the use of "credible" (real) data from an SDR such as DTCC.

We recommend that the OFR work with a credible third party to carry out the project. This is a way to bring additional resources to bear and to the get independent judgment on what's desirable and practical. One option would be to engage MITRE Corporation to carry out the project.[2] We further recommend that the OFR work with the independent third party to establish objective evaluation criteria based on the above test goals in advance of the third party carrying out the demonstration.

We believe that the proposed project can be completed in a timely way because most of the components of the proposed project are already in an advanced state. The project could utilize:

1. **A business conceptual ontology**. One option would be to use the Financial Industry Business Ontology (FIBO) – an open standard created by the EDM Council under the technical governance of the Object Management Group. The business conceptual ontology for both IR swaps and business entities already exists. The project could ensure the completeness of the BCO for IR swaps by evaluating FIBO against the contractual structure of FpML and other relevant data taxonomies.

2. **Mapping of source(s) to the operational ontology**. FIBO has already been mapped to FpML, one of the primary frameworks in which data is already being reported to SDRs. It would probably make sense to verify the FIBO-FpML mapping and extend the mapping process to

[2] The MITRE Corporation is a not-for-profit company that operates multiple federally funded research and development centers (FFRDC). FFRDCs work on projects sponsored by government agencies. What sets FFRDCs apart is their freedom from conflicts of interest—they don't profit from the government's decisions – and they are not allowed to manufacture products, compete with industry, or work for commercial companies. These restrictions mean government and industry can provide FFRDCs with sensitive or proprietary information without fear of improper use or disclosure. This allows FFRDCs to help their government sponsors acquire the right technology, objectively assess business processes, and integrate complex systems.

include sample sets of data from FIXML, ISO 20022, relational databases or CSV files.

3. **Credible (real) data at scale**. We believe that some credible sources will be willing to participate in the project. One issue that will have to be addressed is access to the data. This could be solved by having the CFTC provide authorization to DTCC (or other SDR) to provide data for this purpose, specifying what data the SDR should provide and to whom.

In conclusion, we recommend that the OFR engage a trusted third party to carry out a project to assess the utility of existing ontologies. This project should serve the dual purpose of: (1) helping the OFR demonstrate viable solutions to existing data challenges associated with derivative reporting; and (2) evaluating the potential of ontology to be part of the answer to OFR's instrument database mandate.